CAPTURING AND MODELING KNOWLEDGE OBJECTS: THE SACOT PROJECT

Alain Auger Ph. D.
Defence Research and Development Canada – Valcartier
2459, Pie-XI Blvd North, Val-Belair
Quebec, CANADA G3J 1X5
Alain.Auger@drdc-rddc.gc.ca

1. INTRODUCTION

One of the strategic objectives for Information and Knowledge Management (IKM) in Canadian Command and Control Information Systems (C2IS) consists in investigating and advancing knowledge creation and discovery techniques through which information is collected and processed to support situation analysis and gain sufficient situational awareness to be able to project possible future courses of action or trends with confidence. In 2001, the Future Army Capabilities report (DND, 2001) pointed out that "without some fundamental change, current army ISR[1] will be incapable of providing the degree of knowledge that will be required by future commanders." Therefore "all relevant data, information and knowledge must be available at all levels, but managed in a way that produces a current, rapid and coherent understanding of the battlespace, while at the same time allowing the various levels of command to process the relevant material for their specific purposes."

Initiated in 2004 at Defence Research and Development Canada (DRDC), the SACOT[2] knowledge engineering research project aims at investigating, developing and validating innovative natural language processing (NLP) approaches as scientific means to generate rapidly broad ontologies using large amount of electronic texts.

Ontologies are key elements required to enable decision support systems, knowledge exploitation and information retrieval systems with new semantic capabilities. Since Gruber (Gruber, 1993), the scientific community defines an ontology as a formal, explicit specification of a shared conceptualization. When the knowledge of a domain is represented in a declarative formalism such as in an ontology, the set of objects that can be represented is called the universe of discourse. This set of objects, and the formalized relationships among them, are reflected in the representational *vocabulary* (id.). Domain ontologies provide vocabularies about the concepts within a domain and their relationships, about activities that take place in that domain and about theories and elementary principles governing that domain (Corcho *et al.*, 2003).

[1] Intelligence, Surveillance, Reconnaissance (ISR)
[2] Semi-Automatic Construction of Ontologies from Texts (SACOT)

2. NATURAL LANGUAGE PROCESSING FOR ONTOLOGY ENGINEERING

Knowledge objects of a given domain are expressed and conveyed in texts using domain-specific terminology. As two of the core components of domain ontologies are concepts and relations among concepts, the SACOT project will focus on corresponding NLP research areas. Innovative terminology extraction techniques will be developed, validated and used to automatically build the inventory of concepts of a given domain from representative texts. Semantic relations extraction strategies will be developed and then deployed to identify automatically semantic relations among the concepts of a domain. Draft ontologies will be automatically compiled and generated using the two core components extracted from the corpus (concepts and semantic relations) and other reference material. Knowledge engineers will use this ontology acquisition prototype as a knowledge framework in order to validate and enhance the draft ontologies. Their work will be monitored and supported by machine learning techniques in order to capture validation rules and to provide the system with self-learning capabilities. While validating the content of the draft ontologies, knowledge engineers will teach the system for which of all the potential semantic relations identified in texts are most valuable and which should simply be set aside.

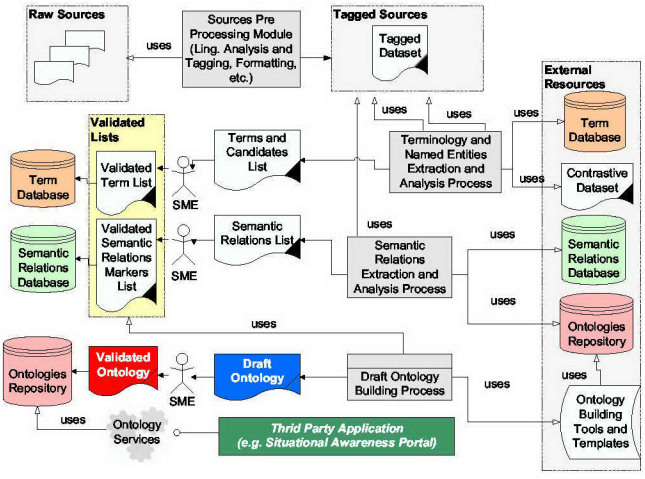

Fig. 1 The SACOT Knowledge Engineering Framework

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
00 DEC 2004	**N/A**	**-**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Capturing and Modeling Knowledge Objectives: The Sacot Project	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Defence Research and Development Canada - Valcartier 2459, Pie-XI Blvd North, Val-Belair Quebec, Canada**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM001736, Proceedings for the Army Science Conference (24th) Held on 29 November - 2 December 2005 in Orlando, Florida. , The original document contains color images.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**UU**	**2**	

3. ONTOLOGY ENGINEERING METHODS

Most of published knowledge engineering methods (Corcho *et al.*, 2003) require interviews with Subject Matter Experts (SMEs). In all the approaches relying heavily on SMEs, the extent of the domain represented in the ontology is, to a certain extent, tailored to the expertise and the degree of "expressiveness" of the SMEs. This limitation might lead to unacceptable and mediocre performance of ontology-based information systems. Typically, domain terminology can contain from few hundreds (e.g. PGA Glossary of Golf) to several hundreds of thousands terms (e.g. up to 160,000 terms in a medical dictionary). It is unlikely that any SME interview will ever elicit the whole terminology of a domain. We need to turn to more exhaustive and objective data sources. The main impediments to classic development of ontologies are that it is a time and budget consuming operation and that it is bound to SMEs' own knowledge limitations. Exhaustive elicitation of knowledge objects of a domain requires the application of NLP extraction techniques over textual data.

4. ONTOLOGIES FOR INTELLIGENT COMMAND AND CONTROL SYSTEMS

Research from the military community (Bourry-Brisset, 2000; Chance *et al.*, 2003; Gauvin *et al.*, 2004, Gouin *et al.*, 2003) clearly indicates that there are many needs and many potential uses for ontologies within the military domain similar to industry. With the development and maturity of the Semantic Web (Davies *et al.*, 2003), the ontology will be the cornerstone technology, which shares a common understanding of a domain among humans, agents and machines. Ontologies for Command and Control Systems will be instrumental in establishing a Common Operational Picture (COP) among units by making domain analysis, situation analysis and assumptions more explicit. Agents assisting commanders with the Command and Control task will have the ability to "interpret" data and know its meaning and value based on the domain ontology.

5. CONCLUSION

SACOT's methods and tools will significantly reduce time usually required to capture the knowledge objects of a domain in traditional ontology building processes. They will provide knowledge engineers with means to rapidly build broad domain ontologies for their applications. Outcomes of this knowledge engineering framework will provide benefits for situational awareness portals, for ontology-based automatic document classification systems, for ontology-based data mining, for knowledge portals, for intelligent search engines and for any other application requiring ontologies to reach semantic-level capabilities. This research project will also be a building block in the development of generic ontology services within the Context – Ontology – Portfolio knowledge management portal project (Gauvin *et al.*, 2004) currently ongoing at DRDC Valcartier.

6. REFERENCES

1. Bourry-Brisset Anne-Claire, 2000: Knowledge Modeling and Management for Command and Control Environments. *Command and Control Research and Technology Symposium (CCRTS)*, June 26-28, Montery, CA.
2. Chance Samuel G., Hagenston Marty G., 2003: *Assessing the Potential Value of Semantic Web Technologies in Support of Military Operations*. Thesis. Naval Postgraduate School. Monterey California United States Navy. 289 p.
3. Corcho Oscar, Fernández-Lopez Mariano Gómez-Pérez Asunción, 2003: Methodologies, Tools and Languages for Building Ontologies. Where Is Their Meeting Point? in *Data and Knowledge Engineering*, **46**, 41-64.
4. Davies John, Fensel Dieter, and Van Harmelen Frank (eds), 2003: *Towards the Semantic Web. Ontology-Driven Knowledge Management.* John Wiley and Sons, LTD ed. England.
5. DND, 2001: Future Army Capabilities in *DLSC Report 01/01*.
6. Gauvin Marlène, Bourry-Brisset Anne-Claire and Auger Alain, 2004 : Context, Ontology and Portfolio: Key Concepts for a Situational Awareness Knowledge Portal in *Proceedings of HICSS-37. Hawaii's International Conference on System Sciences*.
7. Gouin Denis, Gauvin Marlène, and Woodliffe Elizabeth, 2003: COP 21 TD - Towards a Situational Awareness Knowledge Portal in *Proceedings of SPIE 2003 – Aerosense/ Defence Sensing, Simulation and Controls*, Orlando, 21-25 April 2003 Vol. 5101 - Battlespace Digitization and Network-Centric Systems III.
8. Gruber T., 1993: A Translation Approach to Portable Ontologies in *Knowledge Acquisition* **5.2** 199-220.